DOCUMENT RESUME

ED 203 465 EA 013 653

AUTHOR Filipczak, James
TITLE Evaluations of Classroom Observations (ECO). Project Report.
INSTITUTION Institute for Behavioral Research, Silver Spring, Md.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Feb 81
GRANT NIE-G-78-0070
NOTE 37p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavior Change; Behavior Modification; Behavior Problems; *Classroom Observation Techniques; Classroom Research; Junior High Schools; *Research Methodology; Student Behavior; Test Reliability
IDENTIFIERS Disruptive Behavior; Evaluations of Classroom Observations Project

ABSTRACT

The Evaluations of Classroom Observations Project had three major purposes, all related to direct observation of classroom behavior. First, the project was meant to assess the effectiveness of a behavioral treatment program in furthering the acquisition of appropriate classroom participation skills by disruptive students. Second, the project was intended to analyze the effects of teachers, classroom settings, and instructional modes on the physical and verbal behavior of both disruptive and nondisruptive students. Finally, the project was supposed to measure the maintenance of observational skills by trained behavior observers and to examine how observer pairing, scheduling, and placement in certain classes affected the application of these skills. This report describes the methods used to test the reliability of observer measurements, and to establish a control group of students with which to compare those in the behavioral treatment program under study. The students studied were seventh and eighth grade students involved in a program for highly disruptive and skill deficient adolescents in a suburb of Washington, D.C. Behaviors were observed during science classes. Discussion of the data generated through this process is not included in this report, but will be forthcoming pending statistical analysis financed by other funding sources. (PGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED203465

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PROJECT REPORT

Evaluations of Classroom Observations
(ECO)

NIE-G-78-0070

Submitted to: National Institute of Education

Submitted by: James Filipczak
Institute for Behavioral Research
2429 Linden Lane
Silver Spring, MD 20910

February, 1981

CS 013 653

Abstract

The Evaluations of Classroom Observations (ECO) Project intended to assess and suggest refinements for behavioral observation procedures that have been used in carrying out comparative outcomes evaluations. ECO's overall objectives were: 1) to demonstrate the utility of generalizability theory and procedures to outcome evaluation research; 2) to determine if these types of observational procedures were effective in producing reliable data; 3) to examine the influence of instructional setting and occurrence sequence on the behavior of different groups of students; 4) to demonstrate the utility of collecting data on normative peers in the evaluation of a treatment program; and 5) to examine the treatment effectiveness of the original treatment program. Data were available on experimental and control groups from a large-scale behavioral intervention in a local public school. These data included observations of several categories of verbal and physical behavior of students and teachers in a number of science classrooms over one school quarter. In addition, information had also been collected on a range of environmental and treatment variables, including six types of instructional format. The reliability of these original observations had been controlled through a rotated, multiply-blind, and master calibrated set of observers for each of the classroom data collection cycles. All original data were collected from project files, recoded, and reduced to machine readable format for data processing. Behavior frequencies under each of the classroom environment and teacher variables were computed to permit specification of subsequent analytic procedures. The final steps of the reliability and generalizability analyses are to be completed through other funding.

PURPOSES/OBJECTIVES

The objective of the "Evaluations of Classroom Observations" (ECO) Project was to assess and suggest refinements for appraising, interpreting, and planning behavioral observation procedures useful in conducting comparative outcome evaluations.

Research and service personnel frequently use behavioral observation techniques as methods of evaluation within schools. Often, these techniques are selected because they permit direct assessment of student or teacher performance, development, and change where indirect ratings or self-reports do not. Observation techniques are close to the raw material of performance that has been recently seen as most important to assess in education.

The ECO Project had three major purposes, all related to direct observation of intraclassroom behavior:

1. To assess the extent to which disruptive students acquire skills for appropriate classroom participation as a function of a behavioral treatment program.

2. To assess how the physical and verbal behavior of both disruptive and non-disruptive students vary as a function of teacher, classroom setting, and instructional mode.

3. To determine the extent to which trained behavior observers maintain their observational skills and to examine how these skills change as a function of pairing with other observers, placement in certain classes, or scheduling throughout the day.

Each of these purposes has important practical and theoretical significance. Increasing costs of treatment programs and program evaluations necessitate the development and use of assessment procedures which are cost-effective. Yet, researchers using direct observation in behavior assessment

must, at whatever cost, continue to concern themselves with the reliability and validity (including generalizability) of their measure.

Observer Reliability

Although measuring the reliability of instruments has often been a major part of research requiring the use of standardized tests, investigators who carry out observational studies have in many cases failed to assess the reliability of the observers they employ (Johnson & Bolstad, 1973). There may be several sources of observer bias. For example, observers are often aware of the nature and purpose of the research. They may also know that they are being watched, and they may have been told of the expected outcomes. Until these confounding influences have been controlled for and/or assessed, data from observational studies will be difficult to interpret.

The ECO Project demonstrates how use of blind assignment of observers to classroom and students, rotation of observers, and recording of environmental, as well as behavioral events may reduce observer bias, and thus contribute to reliable measurement.

In addition to building safeguards into the experiment itself, staff chose to use several different measures in analyzing observer performance. Each of the coefficients reported in this study and described below yields some unique information about the observers; together, they present a well-rounded picture of the quality and dependability of the observers who were employed.

Interobserver Agreement Percentage. This coefficient expresses the percentage of time units two observers agree about observed behavior and/or setting. This is a useful indicator of observer reliability and is commonly reported in the literature. However, staff did not think this

alone was an adequate way to measure how skillful observers were because this percentage does not provide information about the variability of subjects. Two sets of observations recorded in a classroom in which very little interaction occurs may yield a high interobserver agreement percentage, but this may also indicate that the measure does not detect real, difficult-to-observe differences among subjects. Conversely, observations recorded in a highly active classroom disagree to some extent, but at the same time their observations may distinguish well among subjects.

Interobserver reliability coefficients--sets of correlations between two sets of scores for the same student on the same occasion. Unlike the interobserver agreement percentage, these correlations do reflect real differences among subjects; the error term implied not only contains random error, but also error resulting from differences among observers, be it in the use of the instrument, or the degree of skill of the observers. The size of the error also reflects the complexity of the behaviors observed.

Since both of these coefficients confound random subject error with differences within and between observers, staff decided to carry out a generalizability analysis in the manner of Cronbach et al (1972) in order to further describe observer agreement and identify separate sources of error. Generalizability theory offers a sophisticated approach to reliability and validity questions, in which the investigator identifies a number of "universes of generalization" to which scores or observations can belong. For example, if one could conceptualize a universe of all possible observers, or of all possible occasions or settings, the investigator's objective would be to determine the extent to which these universes contribute to score variance. Sources of variation in a given analysis are identified as facets. All facets do not have to be included in each

analysis--some may be held constant while others are being examined.

Cronbach and his associates postulated two sorts of generalizability analysis. A "G" study is an exploratory look at the amount of variance contributing to total variance for each of the facets used in the particular study. Results of a "G" study can then be used in a "D" or decision study, in which given facets may be manipulated in such a way as to lower their contribution to experimental error.

Subjects

The establishment of a normative comparison base has been a recent concern in the evaluation of treatment programs; a major question in this study was the extent to which the behavior of students who participated in a behavioral treatment program approximated the behavior of non-treated "normal" peers.

PREP (Preparation through Responsive Educational Programs), the behavioral treatment program, was designed for highly disruptive and skill-deficient adolescents. Support by NIMH Grant #MH21950-01, PREP's training program encompassed one school year for each student who was selected and voluntarily participated. Students were recommended for PREP by school staff on the basis of strong evidence of academic and/or social problems: failing grades, high number of disciplinary referrals, chronic absenteeism, frequent tardiness or suspension, poor performance on standardized tests or teacher observations of poor classroom performance, contacts with the police or other juvenile problem agencies, and so forth. Thereafter, staff assessed each student's skills and problems in reference to specific entrance criteria, solicited informed consent of parent and child to participate, developed experimental and control groups as required for research purposes, and enrolled students in the program for that school year.

The program developed for PREP and applied in the school included four distinct but interrelated components:

- Academic training in reading, English language, and mathematics that permitted rapid and thorough skill development.
- Social or interpersonal skill training that facilitated immediate and generalizable behavioral skill development for both in- and out-of-school problems.
- Family liaison and skills training that promoted increased involvement of the parents in school activities and reasonable programs of family management in the home.
- Teacher and other staff training that helped them conduct all phases of the program.

PREP's evaluative procedures began at the time students were enrolled. Data from the criterion-referenced selection procedures formed a major part of the baseline information on each student. PREP used such information to assess the change in student performance from before entrance over the program year on measures like grades, office referrals (for reasons of discipline), suspensions, and attendance. PREP also administered standardized achievement and/or diagnostic tests, the results of which were compared on a pre- and post-program basis to assess academic skill change. A number of personality, attitudinal, and behavioral measures were also taken for such program evaluations. Comparisons were made on these outcome measures between experimental and control groups or across multiple experimental groups, permitting valid determination of the program's effects.

PREP's training program of academic, social, and family skills was comprised of a number of separate but interrelated elements. In certain

years, students assigned to PREP took either all of the training elements or none of the elements, yielding pure experimental and control groups. In other years, some students took none, one, or more of the elements, yielding analyses of different component effects.

The majority of student activity each day was spent in a "skills center" where all academic training was scheduled for a class size of between 16 and 24 students. The students received training in English and mathematics. Most academic learning was managed in a self-instructional medium, with a number of specific topics or objectives organized for small-group instruction. Public school teachers conducted all such programs after receiving training from PREP project staff. Following diagnosis and prescriptive placement, students worked at their own pace on these instructional programs. Learning programs were geared for all major topics and skill levels across a continuum of approximately 2nd through 11th grade objectives. The students' learning progress was assessed continually, through curriculum-embedded tests and with parallel periodic tests. Data were maintained in the program about the students' rate of work, their on-task activities, and their social behavior (according to a number of criteria). These data formed the basis for student ratings by the teachers and eventually determined grades and others means of recognition. Teachers were also trained to create a highly reinforcing atmosphere in the classroom by praising students for their immediate achievements. Yet, they provided a structured atmosphere and practiced assertive teaching procedures to insure that all classroom rules were clearly defined and instruction could be conducted without major disruption.

A limited number of social skills were also trained in these academic classes, principally concerning objectives for self-management of study behavior and individual interaction with other students in a self-study setting.

During a different class period, students participated in a more elaborate social skills training element, often entitled IPS (Interpersonal Skills) class. Here, a large set of social behavior objectives were related to current student skills, and the teacher assisted the students in learning those skills needed for survival in the school, home, and community environments. Basic topics included non-verbal behavior training (such as facial expressions and gestures); small-group interaction, leadership, and followship; personal behavior recording and management; following instructions (in oral, written, and visual forms); and test-taking behavior requirements.

Both small- and large-group instructional processes predominated in this IPS class. Students occasionally worked on individually prescribed units, but most frequently work on a group basis was handled by a single teacher with the assistance of student aides (from outside the class and inside--using better skilled class members). Student behavior was rated and recorded for both content and behavior acquisition. Like the academic classes, students also earned praise from their teacher and the aides when they performed well, earned grades and school credits for skill building, and received other reinforcers for specified achievement. The entire class was run on a behavioral system, with appropriate behavior rewarded and inappropriate behavior ignored or systematically eliminated.

Students were also involved in regular school classes. These were usually limited to four or five other classes per day--including science,

physical education, music or art, social studies, or others as routinely scheduled by the school administration. PREP students participated in these classes to help insure that they were not significantly labeled by participation in PREP and to provide them with other critical skills not a part of PREP training. Yet, these other classes also provided PREP with a means of assessing the degree to which the skills learned in PREP carried over to other settings--an important measure of PREP's success. In some cases, students carried to each of these classes a daily form on which their other class teachers rated their academic and social skills. These forms were returned to PREP by the students, and the data from them formed the basis for routine feedback to the parents. Teachers from these other classes found such daily evaluation to be an important means of providing the student with the additional feedback and control that was often needed. A typical day for a PREP student is illustrated in Figure 1.

Insert Figure 1 about here

Family liaison and training took many forms over the various years of the program. One of the more important features of this program was the "other class" rating described above. However, more substantial elements were devised and used as well. Programs of parent training formed a part of PREP's program in the school. They were conducted by both PREP and school staff and focussed on the development of parent skills in supporting and maintaining student behavior in the home as it pertains to school- and community-related activities. This training was given within the school, during evening hours, along with other community-related training programs. In addition, staff helped the parents develop and conduct

programs of individualized behavior management with their children, often working in concert with the IPS class where the students learned comparable skills, all with the consent of both students and parents. Staff worked with the parents and students in their homes as well, assisting them in the specific conduct of programs of mutual interest and need.

Mindful of the importance of assessing the effects of the PREP Project and evaluating effects of variables such as instructional setting, teachers, and the measuring instruments (the observers) themselves, ECO staff focused on the following specific objectives:

1. To demonstrate the utility of generalizability theory to outcome evaluation research.
2. To ascertain if the rotational and blind observational procedures were effective.
3. To determine the influence of instructional setting and occasion on the behavior of different groups of students.
4. To demonstrate the utility of gathering data on normative peers in the evaluation of treatment programs.
5. To examine the treatment effectiveness of the PREP program.
6. In general, to demonstrate a number of model procedures for evaluating programs.

Data Base/Design of Original Study

The data for this study were gathered in 1973 under an NIMH General Research Support (GRS) grant (No. RR05636-07).

Observations took place in seven different 7th or 8th grade science classrooms in a public junior high school of 1,250 student capacity in the near-city suburbns of Washington, DC. Seven different public school teachers (with no special training or aptitude) were involved, some of

them teaching more than one science period, making a total of 30 different science classes. (Teacher 8 in this study is a composite substitute teacher.)

Observational data were obtained on 24 PREP treatment experimentals; 25 PREP treatment controls, matched on behavior problem and academic deficiency criteria; 162 randomly-selected non-PREP discipline problems; and 83 "normal students" of comparable grade, age, and sex. Treatment and control samples were 70% male and 30% female.

One hundred and seventy days of classroom data were gathered by nine different observers over 25 school days. The observers were trained to 95% or better agreement on the observational task before they were assigned to classrooms. Two observers served as "master calibrators" to provide simultaneous records of observations against which the data obtained by pairs of observers in each classroom were compared.

The observational methodology was based on rigorous, blind assignment of observers to classrooms and students, and across students observed within each of the classrooms. Each of the PREP experimental and control group students (totaling 49) was randomly scheduled to be observed on a given number of different days during the term of the study, and each of the "discipline problems" and "normal" students was scheduled for observation by a random draw of names.

During any one class period, on any one day, one pair of observers recorded classroom setting and behavior. Although 8 students were observed during the class period, 4 of them were observed by one observer only (2 per observer). Two were observed by both observers, but at different times during the class period, and two were observed by both observers simultaneously.

Observers were randomly assigned to conduct approximately four class observations per day, and were given their observation assignment and data-gathering packet just before the class was to start. (Twenty-seven ob erver-pairs took part in the study). Observers were never informed of the group identification of any student, nor were they told of the planned observer overlap on four of the eight students. So not only were they unaware of whether the student was a PREP, a control, a non-problem, or another problem, but they also did not know when interobserver reliability was being assessed.

Upon entering a classroom, the pair of observers sat on stools 20 feet apart at the rear of the classroom, permitting unimpeded observation of all students. Each wore sunglasses to prevent prompting of one another, or "checking each other out" on one behavior or another. The observers began recording behaviors and settings after the class had settled down. (See Appendix for sample of the rating Form 1.)

Each observer observed one assigned student for 10 seconds, recorded physical and verbal behaviors for that interval for 5 seconds, cycled through five remaining assigned students in this same way (for a total of one minute and 30 seconds), rested for 30 seconds, and recycled to student #1 again. During the time that physical and verbal behavior were being recorded, observers also recorded the instructional setting.

Use of a two-channel audio signaling timing device assured synchronization. It was never necessary for the two observers to interact in the classroom. At the end of the class period, each observer independently returned the completed observation record to an office unaffiliated with the other PREP project efforts.

Data collected by the observers were computer-tabulated and summarized by class and observer day (of which there were a total of 170). Though the observation period spanned only four weeks, several observer pairs worked in different classes on any given day. Lack of funds prevented staff from carrying out appropriate statistical analyses. The 170 printouts contained the following information used in the ECO study:

I. Classroom Setting: Observer Agreement. The interobserver agreement percentage for each observer day. Observers recorded in six different categories:

A. Teacher Directed (TD): Lecture, instruction, written work, or tests under teacher supervision, teacher-led discussion.

B. Teacher Directed Individual Work (TDI): Set work.

C. Media Directed (MD): Film; used whenever sound was produced by media, coded only when film had begun.

D. Student Directed (SD): Small group work.

E. Transition Period (TD): Before a film, end of period, teacher verbalized to modify student behavior.

F. Unable to Categorize Setting (X).

Visual inspection of the data led to the dropping of two of these categories in the analysis; neither SD nor X was used more than a few times.

II. Motor Behavior

A. Observer Agreement. The interobserver agreement percentage for each observer day. Observers recorded in 10 different categories.

1. On Task (+): Student was

a. Out of his/her seat and doing something that was task-oriented (e.g., sharpening pencil, getting note from teacher, etc.).

b. Listening--even when student was listening to teacher bawl him/her out.

c. Reading.

d. Writing.

e. On a teacher-designated errand (including going to the lavatory with teacher's permission).

This category was also used when a student was getting an assignment or making a request to leave the romm, and for returning to the room, as long as the student's behavior was appropriate to the occasion. When the student was seemingly attentive (e.g., listening while crumpling paper or playing with a pencil, fidgeting, etc.), observers were instructed to give him/her the benefit of the doubt and record the behavior as +.

2. On Task Actively (++): Student was

a. Reciting.

b. Raising hand.

c. Going to the board.

d. Engaged in group discussion when it was obvious that he/she was discussing work.

3. Off Task (-): Student was

a. Obviously withdrawn (e.g., eyes focused out window).

b. doing homework for other classes.

c. Whispering.

d. Continual fumbling in desk or purse (with eyes on fumbling.

e. Note passing.

f. Talking with teacher off task.

g. Leaving room without designated reason.

h. Picking something up off the floor.

i. Out of his/her set in the room, wandering around.

4. Off Task Actively (--): Student was

a. Talking loudly enough to be audible across room.

b. Engaged in inappropriate physical behavior involving at least one other person (e.g., poking, punching, fiddling with another's hair).

c. Talking loudly enough to teacher to be audible across room.

5. Both On Task and Off Task in Same Interval (+-): Student was

a. Attentive during most of the interval but turned to whisper before the buzzer sounded.

b. Writing or listening, then received or passed a note.

c. On task through most of the interval, but bent down to retrieve something from the floor before the buzzer sounded.

d. On task, but just before the buzzer sounded got up and left the room.

e. Talking to peer, but about class or homework.

f. On task but with inappropriate behavior (e.g., listening and chewing gum, if gum chewing was not allowed, or if the gum chewing included gum popping and blowing bubbles).

6. (++--): Student was volunteering and/or reciting and poking neighbor during same interval.

7. (++-): Raising hand/reciting, and whispering in the same interval.

8. (--+): Poking neighbor, playing with another's hair, or engaged in some other off-task activity, then getting on task.

9. (C): Student was cheating; or (O): Student was engaging in a behavior that observer did not understand or did not seem to fit into any of the above categories.

10. (X): Unable to place student in any of the above categories.

When student was out of the room temporarily, whether for acceptable or unacceptable reasons, the observers were instructed not to record anything. They were also instructed to record as + behavior reading library books or quietly writing during a teacher-designated social period. Some conversing with the teacher during such a period was to be recorded ++, while whispering or talking quietly either to oneself or another student was to be recorded -. Inappropriate behavior such as loud talking, punching, etc., was still recorded as --.

B. Individual Student Data. In addition to observer agreement matrices, for each observer day, 12 setting-by-motor-behavior matrices appeared on the printout: one for each of the four students observed by one of the two observers, two for each of the two students observed by both observers but at different times, and two for each of the two students observed by both observers simultaneously. Visual inspection of these matrices revealed that the following four motor behavior categories were used seldom enough to make meaningful analysis impossible: ++--, ++-, --+, and C/O. These matrices contained rates of behavior within each setting/behavior category and percent of total time spent in each setting/behavior category. The matrices also identified the student by his/her code number, gave the number of the match, the group number, and the observer number.

III. Verbal Behavior

A. Observer Agreement. The interobserver agreement percentage for each observer day. Observers recorded in six different categories.

1. T: Verbalization between student and teacher, when it was unclear who initiated it.
2. S: Verbalization between student and peer, when it was unclear who initiated it.
3. TSC: Verbalization between teacher and class, including student.
4. Non: No verbalization during the time observed.

5. TS: Verbalization between teacher and student--initiated by teacher.

6. X: Unable to place observation in any of the above categories.

The verbal interaction data collection procedures was based on the previous work of Ober (1968).

B. Individual Student Data. In addition to the observer agreement matrices, for each observer day 12 setting-by-verbal-behavior matrices appeared on the printout, one for each of the four students observed by one of the two observers, two for each of the students observed by both observers but at different times, and two for each of the two students observed by both observers simultaneously.

PROCEDURES/PRESENT STUDY

Several of the statistical procedures used in this project were described in the introductory section of this report; additional procedures were carried out in accordance with the stated objectives regarding differential behavior among the four groups of students, establishment of normal scores, assessment of the PREP program, and determination of the influence of instructional setting and occasion on the behavior of the different groups. Before any statistical analysis began, staff were faced with the tasks of reorganizing and reducing the data, writing a codebook, and coding the data for submission to the computer. The data reduction/coding process (including problems that were encountered along the way), and the analytic process that followed are described below.

Data Reduction

I. Observer Agreement. Staff developed a summary form onto which percentage of observer agreement for classroom setting and motor and verbal behavior were recorded (see Form 2.)

This form was also used to rearrange the data by observer pairs (and was designated Form 3). This and all other forms were organized by teacher/day. For example, Teacher 1, who taught only during the 3rd period, would have all five days on one sheet.

II. Individual Student Data

A. Observations per student within period and observation day. This form, which listed all students observed during a given teacher/period on a give sheet, gave staff a clear picture of when, by whom, and how many times each student was observed. It served as a valuable reliability check for the motor and verbal behavior data, and for the pair-wise inter-observer agreement data. (See Form 4).

B. Total number of times individual students were observed in each setting in each motor behavior. (See Form 5).

C. Total number of times individual students were observed in each setting in each verbal behavior. (See Form 6).

Both B and C listed the rates of behavior for each student in each setting during each day the student was observed. Since these were an important source of information and presented several problems, they are described below in some detail.

These forms were necessary because, although the printouts presented day-to-day behavior in matrix form, the program was not designed to summarize rates of motor or verbal behavior for individual students across days and settings. The reduction form identified the student by number, match number, and group numbers, and contained 24 boxes, one for each setting/behavior combination. The original plan was to examine the printouts and record daily rates in the boxes, using one sheet for each

teacher/period. Since many students were observed more than once in each setting, and behavior category, the plan was to sum the numbers in each category, erase the single day rates, and record the summed rate in the appropriate box.

Staff hired five college students to complete the motor behavior forms. They were given an overview of the entire project and were then instructed on how to complete the forms. They were told to record information from within all matrices except two: for the two students observed simultaneously during a given observer day, they were told to record information solely from the matrix on the left hand side of the page. The reasoning behind this arbitrary choice was that including both matrices would give the students in question an inflated behavioral score since, for most of the statistical analyses, the summed rates would be averaged by dividing the total rate by the number of days observed, not by the number of discrete times observed.

After the students had completed this task, staff carried out a reliability check. They randomly selected 13 students who had been observed five times. The reason for requiring five days of observation was that this would represent a start toward examining stability of student behavior across days. Two staff members, working independently, repeated the task of summing rates of motor behavior across days and settings for each of the 13 students. Since the purpose for doing this was partially for gathering stability data, they did not sum the rates over days; each form represented only one observation day, instead of the total number of observation days for any one teacher/period.

When the two staff members compared their results, they found they had been 98% accurate in their recording. However, when they compared their results with those obtained by the college students, they found many errors in the students' work, even over the relatively small sample of 13 students: out of 92 possible recordings (summing across the 13 students used), 68 were correct and 24 were incorrect--an error rate of 22%.

As a consequence, project staff were required to completely re-record all of the data from the study. This was accomplished over approximately 2 weeks of time, using labor that was both paid and unpaid by the ECO grant. Thereafter, each of the data sets were verified by multiple inspection by two different staff. An accuracy rate of 97% was found in the re-recordings made by the ECO staff. This permitted the staff to have a high degree of confidence in the accuracy of the basic data.

As a final set of steps, ECO staff transferred these data onto machine-readable format. This entailed the recording of all verified data onto IBR coding sheets and reverifying the resulting forms for accuracy. These coding sheets were used by a reliable keypunching service to compose 80 column machine-readable cards which were again verified by both the service and IBR staff.

DATA ANALYSIS - CURRENT STATUS

These cards were first used to generate a complete data list and the set was dumped to permit a direct inspection of the data fields. These steps provided an accurate data array for the batch-processing data services proposed for use with these data.

The first statistical procedure carried out with the data set was to obtain frequency distributions of behaviors within classrooms, within

students, within settings, and across students. This output served only as a planning guide for each of the other analyses that were to follow. They permitted staff to judge the normality of data collected on each of the behaviors in each of the settings as a means of demonstrating the applicability and suitability of subsequent analyses. Based on these outputs and the original proposal, staff constructed a tentative final plan of analysis for these data.

At this point, the project had nearly reached the end of its term and its funding. A request was made to and approved by the NIE to expend its remaining funds on the final analysis steps over an extended period of time. To conclude these analyses, Dr. James Breiling, project Co-Principal Investigator, arranged for additional computer time and labor to be provided by the Computer Center of the National Institute of Mental Health. Consequently, all data and control cards were transferred to the NIMH where Dr. Breiling worked with technical staff of the Center in composing an efficient and effective procedure for obtaining the reliability and generalizability analyses. All data were transcribed to tapes at the Center.

At that point (summer of 1980), the NIMH faced a severe budgetary limitation for computer operations which put this project in abeyance for a period of time. This problem was made known to the NIE staff. Funds within the NIMH were not made available over the summer and early fall for this purpose.

Concurrently, the Project Director sought and was granted funds from the Biomedical Research Support Grant funds available at the Institute for Behavioral Research to conclude the analyses of these data. These funds are now available.

Thereafter, the Project Director requested the transfer of the NIMH data tapes to IBR so that the processing could be concluded at the American University Computing Center, given the BRS funding from IBR. These tapes were apparently not compatible with the American University equipment and new versions of the tapes are now being made by the NIMH.

When these tapes are received, the IBR will proceed with the processing of the data set that was developed through the ECO grant from the NIE. These last analyses should be completed by the summer of 1981, using the funds allocated by IBR. Thereafter, within the summer, these articles that were proposed at secondary products of this project will be composed and submitted for publication.

REFERENCES

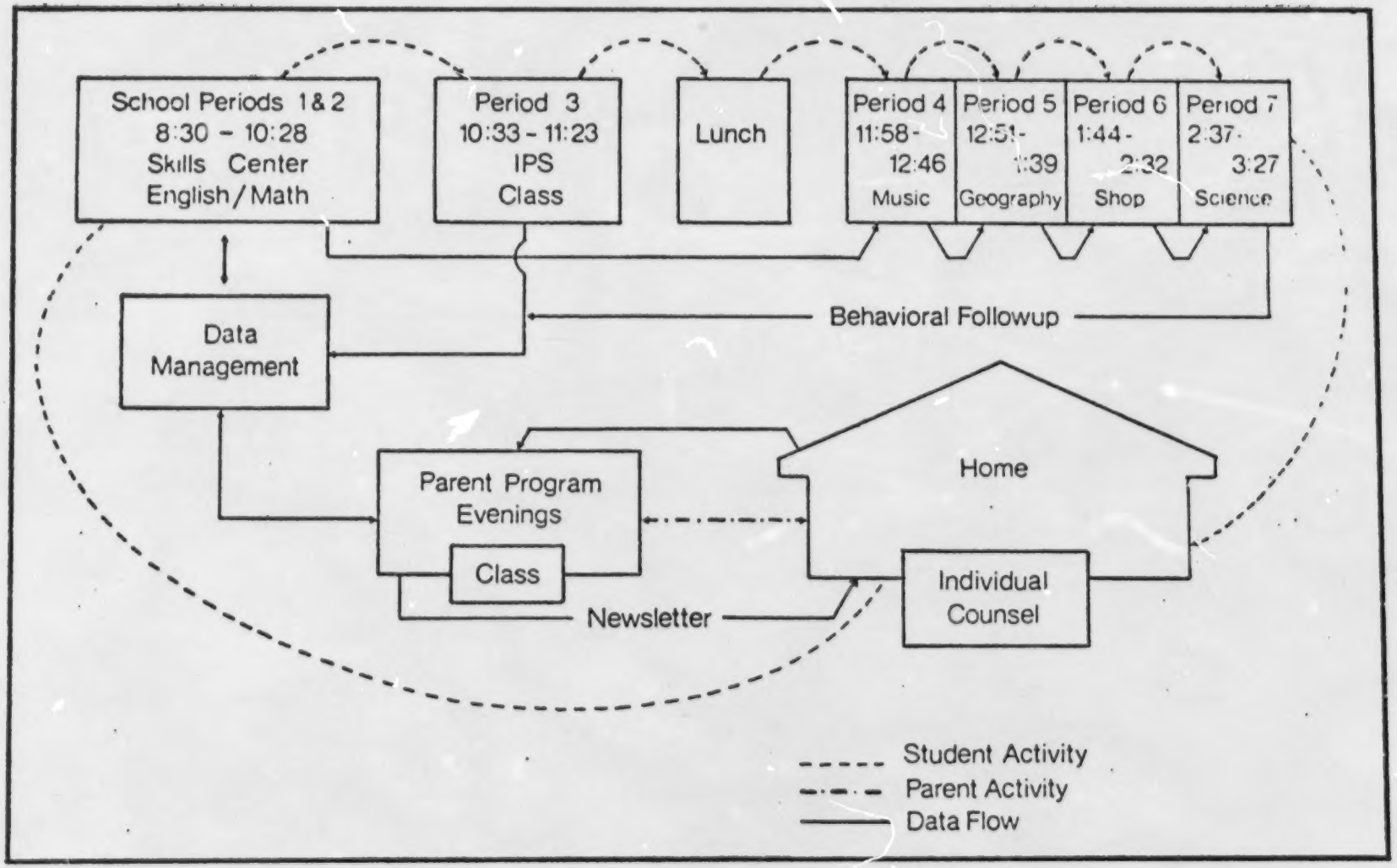
Cronbach, L. J., Gleser, G. C., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley, 1972.

Ober, R. L. The reciprocal category system - RCS. Morgantown, WV: West Virginia University Press, 1968.

FIGURE

FIGURE 1

Example of Daily Activities in the PREP Project



ECO PROJECT FORMS

FORM 1
STUDENT OBSERVER SHEET

Observer _____ Date _____

School _____ Teacher _____

Grade _____ Period _____

1	2	3	4	5	6

Narrative 25 sec. + for recording

1	2	3	4	5	6

Narrative 25 sec. + for recording

1	2	3	4	5	6

Narrative 25 sec. + for recording

1	2	3	4	5	6

Narrative 25 sec. + for recording

1	2	3	4	5	6

Narrative 25 sec. + for recording

1	2	3	4	5	6

Narrative 25 sec. + for recording

1	2	3	4	5	6

Narrative 25 sec. + for recording

FORMS 2 and 3

EVALUATIONS OF CLASSROOM OBSERVATIONS

Observer Agreement--% of A/(A+D)*--for Classroom Setting, Student Motor Behavior, and Student-Teacher Interactions

Teacher: _____

Period: _____

OBSERVATION DAY	CLASSROOM SETTING			STUDENT MOTOR BEHAVIOR			STUDENT-TEACHER VERBAL INTERACTION			
	a	b	c	a	b	c	a	b	d	

*During first half of period (a), second half (b), average (c).

FORM 4

EVALUATIONS OF CLASSROOM OBSERVATIONS

Observations Per Student Within Period and Observation Day

Teacher: _____ Period: _____

STUDENT ID#	MATCH ID#	GROUP ID#	# DAYS OBSERVED/OBSERVER #									TOTAL OBSERVER DAYS PER STUDENT
			1	2	3	4	5	6	7	8	9	
TOTAL OBSERVER DAYS ACROSS STUDENTS												

FORM 5

EVALUATIONS OF CLASSROOM OBSERVATIONS

Teacher: _____

Total Number of Times Individual Students were Observed in Each Setting in Each Motor Behavior

Period: _____

ID#	MATCH #	GROUP #	TD+	TD++	TD-	TD--	TD+-	TDX	MD+	MD++	MD-	MD--	MD+-	MDX	T+	T++	T-	T--	T+-	TX	TDI+	TDI++	TDI-	TDI--	TDI+-	TDIX	TOTAL

FORM 6

EVALUATIONS OF CLASSROOM OBSERVATIONS

Teacher: _____

Total Number of Times Individual Students were Observed in Each Setting in Each Verbal Behavior

Period: _____

ID#	MATCH #	GROUP #	TDS	TDT	TDTSC	TDN	TDTS	TDX	MDS	MDT	MDTSC	MDN	MDTS	MDX	TS	TT	TTSC	TN	TTS	TX	TDIS	TDIT	TDTSC	TDIN	TDITS	TDIX	TOTAL